LETTER

doi:10.1038/nature09869

Genome-wide analysis reveals novel molecular features of mouse recombination hotspots

Fatima Smagulova[1]*, Ivan V. Gregoretti[2]*, Kevin Brick[2], Pavel Khil[2], R. Daniel Camerini-Otero[2] & Galina V. Petukhova[1]

Meiotic recombination predominantly occurs at discrete genomic loci called recombination hotspots, but the features defining these areas are still largely unknown (reviewed in refs 1–5). To allow a comprehensive analysis of hotspot-associated DNA and chromatin characteristics, we developed a direct molecular approach for mapping meiotic DNA double-strand breaks that initiate recombination. Here we present the genome-wide distribution of recombination initiation sites in the mouse genome. Hotspot centres are mapped with approximately 200-nucleotide precision, which allows analysis of the fine structural details of the preferred recombination sites. We determine that hotspots share a centrally distributed consensus motif, possess a nucleotide skew that changes polarity at the centres of hotspots and have an intrinsic preference to be occupied by a nucleosome. Furthermore, we find that the vast majority of recombination initiation sites in mouse males are associated with testis-specific trimethylation of lysine 4 on histone H3 that is distinct from histone H3 lysine 4 trimethylation marks associated with transcription. The recombination map presented here has been derived from a homogeneous mouse population with a defined genetic background and therefore lends itself to extensive future experimental exploration. We note that the mapping technique developed here does not depend on the availability of genetic markers and hence can be easily adapted to other species with complex genomes. Our findings uncover several fundamental features of mammalian recombination hotspots and underline the power of the new recombination map for future studies of genetic recombination, genome stability and evolution.

The vast majority of homologous recombination takes place in recombination hotspots—discrete regions of the genome with a recombination frequency significantly above the frequency in adjacent areas. Extensive studies of several individual hotspots in mammals have greatly advanced our understanding of hotspot biology (reviewed in refs 1–6), but some critical features may be missed without examining the full ensemble of hotspots in a genome. Recently, remarkable progress has been made that culminated in the identification of more than 30,000 recombination hotspots in human[7–11]. Nevertheless, the relatively low resolution of the human map and the high variability of the recombination pattern between individuals still hinders fine structural analysis of hotspots. At the same time, it is clear that the primary DNA structure per se is a poor predictor of recombination activity and that hotspot position depends on additional factors, including epigenetic marks and most probably others[1–4]. Therefore, defining the complete set of variables that ultimately determine hotspot sites in mammals would be greatly facilitated if a hotspot map of a genetically homogeneous and malleable organism were available. We therefore embarked on generating a high-resolution physical map of recombination hotspots in the mouse using a direct molecular approach for identification of recombination initiation sites.

Meiotic recombination is initiated by the introduction of DNA double-stranded breaks (DSBs) by the protein SPO11, followed by resection of the ends to produce long, single-stranded overhangs[12]. Proteins RAD51 and DMC1 form nucleoprotein filaments at the ends of the breaks and search for a homologous chromosome that is used for repair[12]. We used anti-DMC1 antibodies to localize recombination initiation sites in the male mouse genome by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq). To enrich for DSB-stage spermatocytes, we used *Hop2*$^{-/-}$ mice (*Hop2* also known as *Psmc3ip* or *Tbpip*) that lack the cells of later spermatogenic stages owing to meiotic arrest after DSB formation[13] (Supplementary Text). Hotspots identified by our approach in wild-type and in *Hop2*$^{-/-}$ mice correlate extensively (Supplementary Fig. 1), but four times more hotspots can be identified using *Hop2*$^{-/-}$ mice, owing to a higher signal-to-noise ratio.

Consistent with co-localization of RAD51 and DMC1 to DSB sites[12], we found that tag coverage in anti-RAD51 ChIP-Seq was highly correlated with that in anti-DMC1 ChIP-Seq (Fig. 1). Furthermore, the mapping data were highly reproducible between biological replicates ($R = 0.71$–0.97 (correlation coefficient); Fig. 1c). We were able to identify 9,874 recombination hotspots ($P = 10^{-4}$, FDR = 6.7% (false-discovery rate); Supplementary Data and Supplementary Figs 2 and 3), although the number of hotspots could be higher when less restrictive parameters are used (Supplementary Fig. 4). Confirmation of several identified hotspots was carried out by two approaches including direct physical detection of DSBs as previously described[14] (Supplementary Fig. 5). Furthermore, we found that the correlation of our DSB map with available genetic maps[15,16] is almost as high as the correlation between the genetic maps themselves (Fig. 1d, Supplementary Fig. 6 and Supplementary Text). The correlation of these genetic maps with the DSB hotspot distribution from this study is an important validation of our hotspot mapping approach. Additional supporting evidence comes from our finding that the 'hottest' cluster of DSB hotspots in the mouse genome is located in the PAR—the only homologous region between the X and Y chromosomes (Fig. 1e and Supplementary Fig. 3). Despite the very short length of the PAR, each spermatocyte undergoes an obligatory crossover in this area[17]. Estimates from our analysis show that this cluster is probably sufficient to ensure that there is at least one DSB in the PAR of every spermatocyte (Supplementary Text).

The centres of the hotspots in our map are defined with an approximate precision of 200 nucleotides, which is at least an order of magnitude higher than that of other available recombination maps in multicellular organisms (Supplementary Methods). More than 250 of the identified hotspots are hotter than one of the hottest previously known hotspots, *H2Eα* (Fig. 2a). On the basis of the estimated maximum recombination frequency of the *H2Eα* hotspot, 2 cM (ref. 18), the strength of the hottest hotspot in the mouse might be as high as 6 cM. Most of the hotspots lie 60–330 kb apart, with only a few recombination deserts more than 3 Mb in length (Supplementary Fig. 7). The average strengths of individual hotspots on different autosomes are similar, and on a chromosome scale the slight variation in hotspot density between autosomes does not correlate with the density of genes, chromosomal GC content, abundance of DNA repeats or chromosome length (not shown).

[1]Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814, USA. [2]National Institute of Diabetes, Digestive and Kidney Diseases, NIH, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

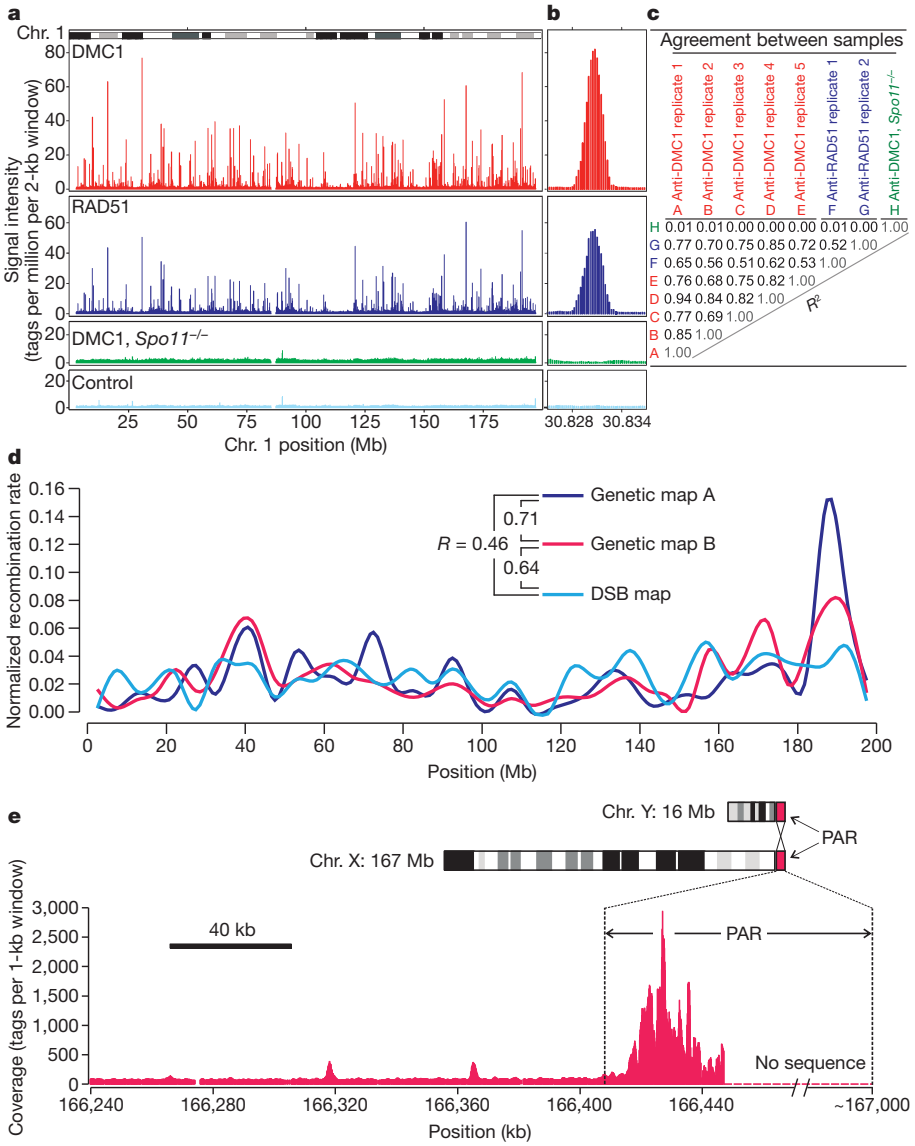

Figure 1 | DSB hotspots in the mouse genome. **a**, ChIP-Seq tag density profiles. DMC1: anti-DMC1 ChIP; RAD51: anti-RAD51 ChIP; DMC1, *Spo11*$^{-/-}$: anti-DMC1 ChIP from *Spo11*$^{-/-}$ mice that do not form DSBs; Control: IgG ChIP and input DNA pool. **b**, Close-up of a representative hotspot. **c**, Agreement between ChIP-Seq samples (correlations in 2-kilobase (kb) bins across genome). **d**, Correlation between the DSB hotspot map and the published genetic maps A[15] and B[16] for chromosome 1. The DSB map is generated from hotspot strengths (Supplementary Methods). All maps are generated in 5-megabase (Mb) windows and normalized by area of the map. **e**, The pseudoautosomal region (PAR) contains a large cluster of overlapping hotspots. DMC1 ChIP-Seq tag coverage is shown (smoothing window, 1 kb; step size, 100 bp).

We found that mouse hotspots have a tendency to overlap genes (Fig. 2b), although only the hottest 40% of the hotspots contribute significantly to this correlation (Supplementary Fig. 8). We next examined the association of DSB hotspots with different classes of DNA repeats and other genomic features. Significant correlations were found with GC content, short and long interspersed nuclear elements, long terminal repeats and other repeats, in agreement with those found in previous work[1,2,4,5] (Supplementary Table 1 and Supplementary Text).

In searching for additional features that might define hotspot locations, we analysed the nucleotide composition of the hotspot regions, as characteristic nucleotide skews have been found at some functional genomic elements including replication origins and transcription start sites[19]. Examination of either single strand of the double-stranded DNA in the 5′-to-3′ direction reveals that the sequence 5′ to the centre of a hotspot is enriched in purines but that the polarity of the bias changes in the middle of the hotspot, such that the sequence 3′ to the hotspot centre is more rich in pyrimidines (Fig. 2c and Supplementary Fig. 9). Replication- and transcription-related skews have been attributed to mutational biases acting asymmetrically on complementary DNA strands. It is conceivable that the skew detected at hotspots is the result of mutational asymmetry as well (Supplementary Text and Supplementary Fig. 9c). It is also possible that the skew represents some unknown functional feature of the genome that favours DSB formation. Importantly, we were able to detect the same signature of nucleotide usage in human hotspots (Supplementary Fig. 9d), suggesting that the purine–pyrimidine skew is an intrinsic property of recombination hotspots in mammals. We also noticed a slight increase in the overall GC content in the middle of hotspots (not shown), which might indicate the presence of a gene conversion bias[20].

The protein PRDM9 is a meiosis-specific methyltransferase responsible for trimethylation of lysine 4 on histone H3[21] (H3K4me3). In addition to its well-known role in transcription[22], this histone modification is associated with increased recombination activity in yeast and in mouse[23,24]. PRDM9 is highly polymorphic in its multi-Zn^{2+}-finger DNA-binding domain, and recent studies have implicated PRDM9 in determining the positions of recombination hotspots through the

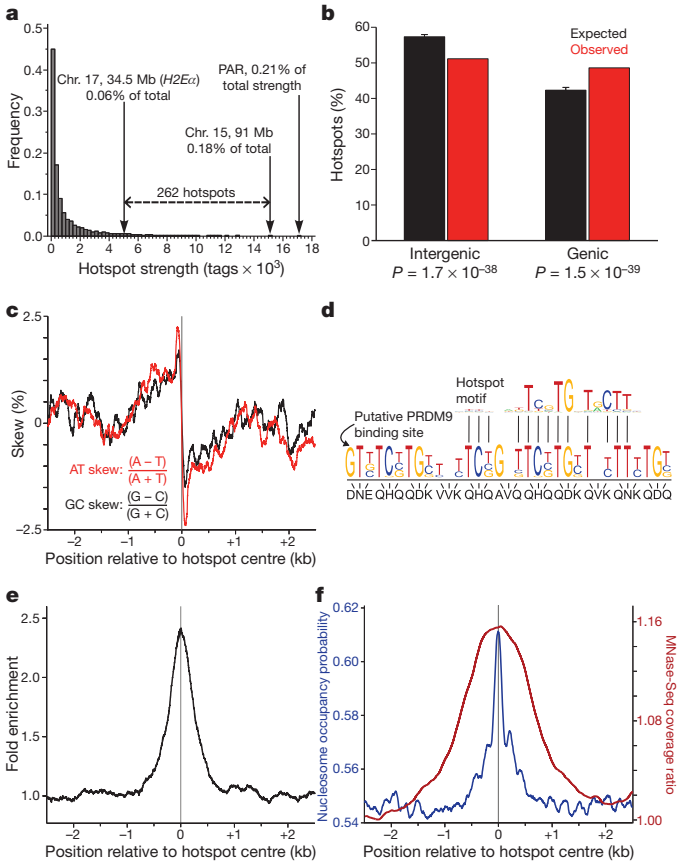

Figure 2 | Characteristics of mouse DSB hotspots. **a**, The *H2Eα* hotspot and the PAR hotspot cluster are among the strongest in the mouse genome. The strength of the hottest individual hotspot identified in this study (chromosome 15, ~91 Mb) is also indicated. **b**, Mouse DSB hotspots are significantly enriched in genes (one-sided binomial tests). Genic regions are defined from start to stop codons including introns. Error bars, 5th to 95th percentiles of the expected value distributions ($n = 10{,}000$ iterations). **c**, A purine–pyrimidine skew is apparent at DSB hotspots. Skew is calculated in 100-bp windows with a step size of 1 bp. **d**, The consensus hotspot motif is similar to the predicted binding site of PRDM9. **e**, The consensus motif is present in the centres of DSB hotspots. The distribution of hits to the consensus motif is shown in the 5-kb regions around hotspots (window size, 200 bp; step size, 1 bp). **f**. Both predicted (blue) and experimentally determined (red) nucleosome occupancy profiles peak at the centres of DSB hotspots. MNase-Seq coverage ratio is plotted as the whole-fragment coverage ratio of micrococcal-nuclease-digested chromatin to randomly fragmented chromatin in sliding, 500-bp windows (step size, 1 bp).

different binding specificities of its alleles[25–27]. Approximately 40% of human recombination hotspots possess a consensus motif[28] that matches the predicted binding site of human PRDM9. We were able to identify a consensus motif specific to mouse hotspots (Fig. 2d, e). Sequences with better alignment scores to the motif consensus are more strongly over-represented in the hotspot regions, with the best hits showing almost 180-fold enrichment (Supplementary Fig. 10). Overall, hotspots containing consensus sequences are stronger than those without, and the quality of the motif alignment within hotspots is positively correlated with hotspot strength (Supplementary Fig. 11). Importantly, the motif shows a strong match to the predicted binding site of the *Prdm9* allele present in our mouse strain (Fig. 2d and Supplementary Fig. 12), and it is present at the centre of at least 73% of hotspots. This indicates that PRDM9 is a determinant for many more hotspots than previously thought and can explain the stronger-than-expected correlation between hotspot activity and *Prdm9* allelic variation found in recent association[27] and sperm typing[29] studies.

Because post-translational histone modification has been implicated in the regulation or maintenance of recombinational activity, we asked whether DNA in the hotspot regions is assembled into nucleosomes. Strikingly, we found that both the predicted[30] and the actual nucleosomal occupancies are co-centred with recombination hotspots (Fig. 2f), reflecting a previously unknown, intrinsic ability of hotspot DNA to assemble a nucleosome. We next examined the distribution of the H3K4me3 marks in germ (testis) and somatic (liver) tissues. We found that 94% of hotspots overlap H3K4me3, with the majority of hotspots overlapping testis-specific H3K4me3 marks (87%) and practically none overlapping the marks specific to the liver (Fig. 3a and Supplementary Fig. 13). Although the association of H3K4me3 with DSB hotspots is not surprising in light of recent work[23–27], the extent of the overlap revealed here is such that H3K4me3 can be considered a global feature of DSB sites in multicellular organisms. Importantly, unlike in *Saccharomyces cerevisiae*[23], where gene promoters and hotspots seem to share the same H3K4me3 mark, H3K4me3 marks at mouse hotspots are hotspot specific (Fig. 3b). Even when a hotspot is located very close to a transcription start site, the corresponding H3K4me3 marks are clearly spatially distinct (Fig. 3c), indicating that different mechanisms are involved in the histone modifications at these sites (Fig. 3c, Supplementary Fig. 13 and Supplementary Text).

Importantly, H3K4me3 per se is not a sufficient mark for DSB formation. Why meiotic DSBs are correlated with only a small fraction (16.7%) of the ~55,000 H3K4me3 marks in testis remains a mystery. An attractive possibility is that some component of the DSB machinery directly associates with PRDM9 or a PRDM9-containing complex and is therefore delivered to the potential DSB sites. Subsequent trimethylation of H3K4 might be required to set the stage for recombination initiation and progression. Other possibilities may also be considered: (i) in addition to H3K4me3, other epigenetic marks may be present that require H3K4me3 to allow DSB formation; (ii) H3K4me3 introduced

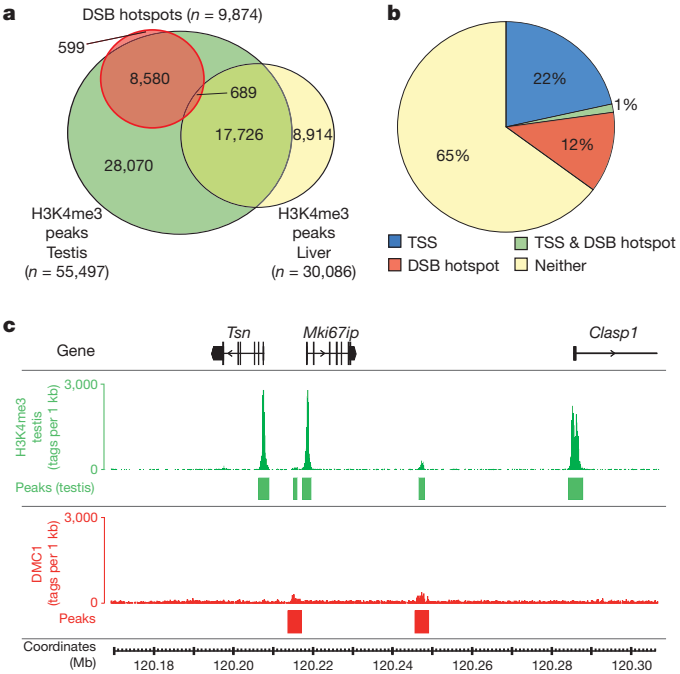

Figure 3 | Specific H3K4me3 marks are associated with DSB hotspots. **a**, The vast majority (93.9%) of DSB hotspots overlap H3K4me3 marks, most of which (86.9%) are testis specific. The six DSB hotspots that overlap liver-specific H3K4me3 marks are not shown. Peak calling for each data set was performed using an equal number of tags. **b**, DSB hotspots are associated with a set of H3K4me3 marks that are distinct from those at transcription start sites (TSSs). The fraction of H3K4me3 marks overlapping TSSs and/or DSB hotspots, or neither, is shown. **c**, H3K4me3 marks at DSB hotspots are generally weaker than TSS-associated marks, and are also spatially distinct despite being sometimes in very close proximity. Tag coverage is displayed in 100-bp steps.

by PRDM9 may be different from other H3K4me3 marks (for example, PRDM9, unlike other methyltransferases, might modify one rather than both H3 histones in the same nucleosome, or the other way around); (iii) specific histone variants could be present in the PRDM9-modified nucleosomes or could be substrates of PRDM9; (iv) transcription factors or other proteins bound to 'non-PRDM9' trimethylation marks may interfere with the DSB machinery.

METHODS SUMMARY

We used *Hop2*$^{-/-}$ mice[13] on a [C57BL/10.S × C57BL/10.F] F_1 genetic background and wild-type mice on the same background to make a map of DSB hotspots. ChIP and high-throughput sequencing were performed according to manufacturer-provided protocols (Upstate and Illumina, respectively) with minor modifications. For each sample, 36-bp end sequences were aligned to the mm9 reference genome using the Illumina GAII analysis pipeline. Only quality-filtered reads that mapped uniquely to the genome were retained for downstream analyses. We identified DSB hotspots and H3K4me3 peaks by comparing the sequence tag coverage for each ChIP sample with that of tag-count-matched control samples using the MACS algorithm. The hotspot consensus motif was identified from nonamers enriched near hotspot centres using a bespoke analysis pipeline. Additional details are available in Supplementary Information.

Received 5 October 2010; accepted 24 January 2011.
Published online 3 April 2011.

1. Arnheim, N., Calabrese, P. & Tiemann-Boege, I. Mammalian meiotic recombination hot spots. *Annu. Rev. Genet.* **41,** 369–399 (2007).
2. Buard, J. & de Massy, B. Playing hide and seek with mammalian meiotic crossover hotspots. *Trends Genet.* **23,** 301–309 (2007).
3. Lichten, M. Meiotic chromatin: the substrate for recombination initiation. *Genome Dynam. Stabil.* **3,** 165–193 (2008).
4. Paigen, K. & Petkov, P. Mammalian recombination hot spots: properties, control and evolution. *Nature Rev. Genet.* **11,** 221–233 (2010).
5. Clark, A. G., Wang, X. & Matise, T. Contrasting methods of quantifying fine structure of human recombination. *Annu. Rev. Genom. Hum. Genet.* **11,** 45–64 (2010).
6. Kauppi, L., May, C. A. & Jeffreys, A. J. Analysis of meiotic recombination products from human sperm. *Methods Mol. Biol.* **557,** 323–355 (2009).
7. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
8. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
9. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).
10. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467,** 1099–1103 (2010).
11. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310,** 321–324 (2005).
12. Neale, M. J. & Keeney, S. Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* **442,** 153–158 (2006).
13. Petukhova, G. V., Romanienko, P. J. & Camerini-Otero, R. D. The Hop2 protein has a direct role in promoting interhomolog interactions during mouse meiosis. *Dev. Cell* **5,** 927–936 (2003).
14. Qin, J., Richardson, L. L., Jasin, M., Handel, M. A. & Arnheim, N. Mouse strains with an active H2-Ea meiotic recombination hot spot exhibit increased levels of H2-Ea-specific DNA breaks in testicular germ cells. *Mol. Cell. Biol.* **24,** 1655–1666 (2004).
15. Paigen, K. *et al.* The recombinational anatomy of a mouse chromosome. *PLoS Genet.* **4,** e1000119 (2008).
16. Cox, A. *et al.* A new standard genetic map for the laboratory mouse. *Genetics* **182,** 1335–1344 (2009).
17. Burgoyne, P. S. Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum. Genet.* **61,** 85–90 (1982).
18. Khambata, S., Mody, J., Modzelewski, A., Heine, D. & Passmore, H. C. Ea recombinational hot spot in the mouse major histocompatibility complex maps to the fourth intron of the Ea gene. *Genome Res.* **6,** 195–201 (1996).
19. Francino, M. P. & Ochman, H. Strand asymmetries in DNA evolution. *Trends Genet.* **13,** 240–245 (1997).
20. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genom. Hum. Genet.* **10,** 285–311 (2009).
21. Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323,** 373–375 (2009).
22. Wang, Z., Schones, D. E. & Zhao, K. Characterization of human epigenomes. *Curr. Opin. Genet. Dev.* **19,** 127–134 (2009).
23. Borde, V. *et al.* Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* **28,** 99–111 (2008).
24. Buard, J., Barthes, P., Grey, C. & de Massy, B. Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *EMBO J.* **28,** 2616–2624 (2009).
25. Parvanov, E. D., Petkov, P. M. & Paigen, K. *Prdm9* controls activation of mammalian recombination hotspots. *Science* **327,** 835 (2010).
26. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327,** 876–879 (2010).
27. Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327,** 836–840 (2010).
28. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genet.* **40,** 1124–1129 (2008).
29. Berg, I. L. *et al. PRDM9* variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genet.* **42,** 859–863 (2010).
30. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458,** 362–366 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Lichten (NCI, NIH) and P. Hsieh (NIDDK, NIH) for comments and discussion. We are grateful to S. Sharmeen for her help with high-throughput sequencing. This work was supported in part by Basil O'Connor Starter Scholar Research Award Grant No. 5-FY07-667 from the March of Dimes Foundation (G.V.P.); NIH grant 1R01GM084104-01A1 from NIGMS (G.V.P.); New Investigator Start-up Grants FS71HU, R071HU and CS71HU from USUHS (G.V.P.); and the NIDDK (NIH) Intramural Research Program (R.D.C.-O.).

Author Contributions F.S. performed all experiments. I.V.G., K.B. and P.K. performed computational data analyses. All authors contributed to experimental design. G.V.P. and R.D.C.-O. designed and supervised the study. G.V.P. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information ChIP-Seq data have been deposited in the Gene Expression Omnibus under accession number GSE24438. *Prdm9* complementary DNA sequences have been deposited in GenBank under accession numbers HQ704390 and HQ704391. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.V.P. (gpetukhova@usuhs.mil) or R.D.C.-O. (rdcamerini@mail.nih.gov).